

SZEGÖ'S THEOREM AND ITS PROBABILISTIC DESCENDANTS

N. H. BINGHAM

Abstract

The theory of orthogonal polynomials on the unit circle (OPUC) dates back to Szegő's work of 1915-21, and has been given a great impetus by the recent work of Simon, in particular his two-volume book [Si4], [Si5], the survey paper (or summary of the book) [Si3], and the book [Si9], whose title we allude to in ours. Simon's motivation comes from spectral theory and analysis. Another major area of application of OPUC comes from probability, statistics, time series and prediction theory; see for instance the book by Grenander and Szegő [GrSz]. Coming to the subject from this background, our aim here is to complement [Si3] by giving some probabilistically motivated results. We also advocate a new definition of long-range dependence.

AMS 2000 subject classifications. Primary 60G10, secondary 60G25.

Key words and phrases. Stationary process, prediction theory, orthogonal polynomials on the unit circle, partial autocorrelation function, moving average, autoregressive, long-range dependence, Hardy space, cepstrum.

CONTENTS

- §1. Introduction
- §2. Verblunsky's theorem and partial autocorrelation
- §3. Weak conditions: Szegő's theorem
- §4. Strong conditions: Baxter's theorem
- §5. Strong conditions: the strong Szegő theorem
- §6. Intermediate conditions
 - 6.1. Complete regularity
 - 6.2. Positive angle: the Helson-Szegő and Helson-Sarason conditions
 - 6.3. Pure minimality
 - 6.4. Rigidity; (LM) , (CND) , (IPF)
- §7. Remarks
- Acknowledgements

References

§1. Introduction

The subject of orthogonal polynomials on the real line (OPRL), at least some of which forms part of the standard undergraduate curriculum, has its roots in the mathematics of the 19th century. The name of Gabor Szegő (1895-1985) is probably best remembered nowadays for two things: co-authorship of 'Pólya and Szegő' [PoSz] and authorship of 'Szegő' [Sz4], his book of 1938, still the standard work on OPRL. Perhaps the key result in OPRL concerns the central role of the three-term recurrence relation ([Sz4], III.3.2: 'Favard's theorem').

Much less well known is the subject of orthogonal polynomials on the unit circle (OPUC), which dates from two papers of Szegő in 1920-21 ([Sz2], [Sz3]), and to which the last chapter of [Sz4] is devoted. Again, the key is the appropriate three-term recurrence relation, the *Szegő recursion* or *Durbin-Levinson algorithm* (§2). This involves a sequence of coefficients (not *two* sequences, as with OPRL), the *Verblunsky coefficients* $\alpha = (\alpha_n)$ (§2), named (there are several other names in use) and systematically exploited in the magisterial two-volume book on OPUC ([Si4], [Si5]) by Barry Simon. See also his survey paper [Si3], written from the point of view of analysis and spectral theory, the survey [GoTo], and his recent book [Si9].

Complementary to this is our own viewpoint, which comes from probability and statistics, specifically *time series* (as does the excellent survey of 1986 by Bloomfield [Bl3]). Here we have a stochastic process (random phenomenon unfolding with time) $X = (X_n)$ with n integer (time discrete, as here, corresponds to compactness of the unit circle by Fourier duality, whence the relevance of OPUC; continuous time is also important, and corresponds to OPRL).

We make a simplifying assumption, and restrict attention to the *stationary* case. The situation is then invariant under the *shift* $n \mapsto n + 1$, which makes available the powerful mathematical machinery of Beurling's work on invariant subspaces ([Beu]; [Nik1]). While this is very convenient mathematically, it is important to realize that this is both a strong restriction and one unlikely to be satisfied exactly in practice. One of the great contributions of the statistician and econometrician Sir Clive Granger (1934-2009) was to demonstrate that statistical/econometric methods appropriate for stationary situations can, when applied indiscriminately to non-stationary situations, lead to misleading conclusions (via the well-known statistical problem

of *spurious regression*). This has profound implications for macroeconomic policy. Governments depend on statisticians and econometricians for advice on interpretation of macroeconomic data. When this advice is misleading and mistaken policy decisions are implemented, avoidable economic losses (in terms of GDP) may result which are large-scale and permanent (cf. Japan's 'lost decade' in the 1990s, or lost two decades, and the global problems of 2007-8 on).

The mathematical machinery needed for OPUC is *function theory on the (unit) disc*, specifically the theory of Hardy spaces and Beurling's theorem (factorization into inner and outer functions and Blaschke products). We shall make free use of this, referring for what we need to standard works (we recommend [Du], [Ho], [Gar], [Koo1], [Nik1], [Nik2]), but giving detailed references. The theory on the disc (whose boundary the circle is compact) corresponds analytically to the theory on the upper half-plane, whose boundary the real line is non-compact (for which see e.g. [DymMcK]). Probabilistically, we work on the disc in discrete time and the half-plane in continuous time. In each case, what dominates is an *integrability condition*. In discrete time, this is *Szegő's condition* (Sz), or *non-determinism* (ND) – integrability of the logarithm $\log w$ of the spectral density w (of μ) (§3). In continuous time, this is the *logarithmic integral*, which gives its name to Koosis' book [Koo2].

In view of the above, the natural context in which to work is that of complex-valued stochastic processes, rather than real-valued ones, in discrete time. We remind the reader that here the Cauchy-Schwarz inequality tells us that correlation coefficients lie in the unit disc, rather than the interval $[-1, 1]$.

The time-series aspects here go back at least as far as the work of Wiener [Wi1] in 1932 on generalized harmonic analysis, GHA (which, incidentally, contains a good historical account of the origins of spectral methods, e.g. in the work of Sir Arthur Schuster in the 1890s on heliophysics). During World War II, the *linear filter* (linearity is intimately linked with Gaussianity) was developed independently by Wiener in the USA [Wi2], motivated by problems of automatic fire control for anti-aircraft artillery, and Kolmogorov in Russia (then USSR) [Kol]. This work was developed by the Ukrainian mathematician M. G. Krein over the period 1945-1985 (see e.g. [Dym]), by Wiener in the 1950s ([Wi3], IG, including commentaries) and by I. A. Ibragimov (1968 on).

The subject of time series is of great practical importance (e.g. in econo-

metrics), but suffered within statistics by being regarded as 'for experts only'. This changed with the 1970 book by Box and Jenkins (see [BoxJeRe]), which popularized the subject by presenting a simplified account (including an easy-to-follow model-fitting and model-checking recipe), based on ARMA models (AR for autoregressive, MA for moving average). The ARMA approach is still important; see e.g. Brockwell and Davis [BroDav] for a modern textbook account. The realization that the Verblunsky coefficients α of OPUC are actually the *partial autocorrelation function* (PACF) of time series opened the way for the systematic exploitation of OPUC within time series by a number of authors. These include Inoue, in a series of papers from 2000 on (see especially [In3] of 2008), and Inoue and Kasahara from 2004 on (see especially [InKa2] of 2006).

Simon's work ([Si3], [Si4], [Si5]) focusses largely on four conditions, two weak (and comparable) and two strong (and non-comparable). Our aim here is to complement the expository account in [Si3] by adding the time-series viewpoint. This necessitates adding (at least) five new conditions. Four of these (comparable) we regard as intermediate, the fifth as strong. In our view, one needs *three* levels of strength here, not two. One is reminded of the Goldilocks principle (from the English children's story: not too hot/hard/high/..., not too cold/soft/low/..., but just right).

We begin in §2 by presenting the basics (Verblunsky's theorem, PACF). We turn in §3 to weak conditions (Szegő's condition (Sz), or (ND); Szegő's theorem; $\alpha \in \ell_2$; $\sigma > 0$). In §4 we look at our first strong condition, Baxter's condition (B), and Baxter's theorem ($\alpha \in \ell_1$). The satisfaction or otherwise of Baxter's condition (B) marks the transition between short- and long-range dependence. The second strong condition, the strong Szegő condition (sSz), follows in §5 (strong Szegő limit theorem, Ibragimov's theorem, Golinskii-Ibragimov theorem, Borodin-Okounkov formula; $\alpha \in H^{1/2}$), together with a weakening of (sSz), absolute regularity. We turn in §6 to intermediate conditions: in decreasing order of strength, (i) complete regularity; (ii) positive angle (Helson-Szegő, Helson-Sarason and Sarason theorems); (iii) (pure) minimality (Kolmogorov); (iv) rigidity (Sarason), Levinson-McKean condition (LM), complete non-determinism (CND), intersection of past and future (IPF); see [KaBi] for details. We close in §7 with some remarks.

The (weak) Szegő limit theorem dates from 1915 [Sz1], the strong Szegő limit theorem from 1952 [Sz5]. Simon ([Si4], 11) rightly says how remarkable it is for one person to have made major contributions to the same area 37 years apart. We note that Szegő's remarkable longevity here is actually

exceeded (over the 40 years 1945-1985) by that of the late, great Mark Grigorievich Krein (1907-1989).

What follows is a survey of this area, which contains (at least) eight different layers, of increasing (or decreasing) generality. This is an increase on Simon's (basic minimum of) four. We hope that no one will be deterred by this increase in dimensionality, and so in apparent complexity. Our aim is the precise opposite: to open up this fascinating area to a broader mathematical public, including the time-series, probabilistic and statistical communities. For this, one needs to open up the 'grey zone' between the strong and weak conditions, and examine the third category, of intermediate conditions. We focus on these three levels of generality. This largely reduces the effective dimensionality to three, which we feel simplifies matters. Mathematics should be made as simple as possible, but not simpler (to adapt Einstein's immortal dictum about physics).

We close by quoting Barry Simon ([Si8], 85): "It's true that until Euclidean Quantum Field Theory changed my tune, I tended to think of probabilists as a priesthood who translated perfectly simple functional analytic ideas into a strange language that merely confused the uninitiated." He continues: in his 1974 book on Euclidean Quantum Field Theory, "the dedication says: "To Ed Nelson who taught me how unnatural it is to view probability theory as unnatural" ".

§2. Verblunsky's theorem and partial autocorrelation.

Let $X = (X_n : n \in \mathbb{Z})$ be a discrete-time, zero-mean, (wide-sense) stationary stochastic process, with autocovariance function $\gamma = (\gamma_n)$,

$$\gamma_n = E[X_n \overline{X_0}]$$

(the variance is constant by stationarity, so we may take it as 1, and then the autocovariance reduces to the autocorrelation).

Let \mathcal{H} be the Hilbert space spanned by $X = (X_n)$ in the L_2 -space of the underlying probability space, with inner product $(X, Y) := E[X \overline{Y}]$ and norm $\|X\| := [E(|X|^2)]^{1/2}$. Write T for the unit circle, the boundary of the unit disc D , parametrised by $z = e^{i\theta}$; unspecified integrals are over T .

Theorem 1 (Kolmogorov Isomorphism Theorem). There is a process Y on T with orthogonal increments and a probability measure μ on T with

(i)

$$X_n = \int e^{int} dY(t);$$

(ii)

$$E[dY(t)^2] = d\mu(t).$$

(iii) The autocorrelation function γ then has the *spectral representation*

$$\gamma_n = \int e^{-in\theta} d\mu(\theta).$$

(iv) One has the *Kolmogorov isomorphism* between \mathcal{H} (the *time domain*) and $L_2(\mu)$ (the *frequency domain*) given by

$$X_t \leftrightarrow e^{it}, \quad (KIT)$$

for integer t (as time is discrete).

Proof. Parts (i), (ii) are the Cramér representation of 1942 ([Cra], [Do] X.4; Cramér and Leadbetter [CraLea] §7.5). Part (iii), due originally to Herglotz in 1911, follows from (i) and (ii) ([Do] X.4, [BroDav] §4.3). Part (iv) is due to Kolmogorov in 1941 [Kol]. All this rests on Stone's theorem of 1932, giving the spectral representation of groups of unitary transformations of linear operators on Hilbert space; see [Do] 636-7 for a historical account and references (including work of Khintchine in 1934 in continuous time), [DunSch] X.5 for background on spectral theory. //

The reader will observe the link between the Kolmogorov Isomorphism Theorem and (ii), and its later counterpart from 1944, the Itô Isomorphism Theorem and $(dB_t)^2 = dt$ in stochastic calculus.

To avoid trivialities, we suppose in what follows that μ is *non-trivial* – has infinite support.

Since for integer t the $e^{it\theta}$ span polynomials in $e^{i\theta}$, prediction theory for stationary processes reduces to approximation by polynomials. This is the classical approach to the main result of the subject, Szegő's theorem (§2 below); see e.g. [GrSz], Ch. 3, [Ach], Addenda, B. We return to this in §7.7 below.

We write

$$d\mu(\theta) = w(\theta)d\theta/2\pi + d\mu_s(\theta),$$

so w is the *spectral density* (w.r.t. normalized Lebesgue measure) and μ_s is the *singular part* of μ .

By stationarity,

$$E[X_m \overline{X_n}] = \gamma_{|m-n|}.$$

The *Toeplitz matrix* for X , or μ , or γ , is

$$\Gamma := (\gamma_{ij}), \quad \text{where} \quad \gamma_{ij} := \gamma_{|i-j|}.$$

It is positive definite.

For $n \in N$, write $\mathcal{H}_{[-n,-1]}$ for the subspace of \mathcal{H} spanned by $\{X_{-n}, \dots, X_{-1}\}$ (the finite past at time 0 of length n), $P_{[-n,-1]}$ for projection onto $\mathcal{H}_{[-n,-1]}$ (thus $P_{[-n,-1]}X_0$ is the best linear predictor of X_0 based on the finite past), $P_{[-n,-1]}^\perp := I - P_{[-n,-1]}$ for the orthogonal projection (thus $P_{[-n,-1]}^\perp X_0 := X_0 - P_{[-n,-1]}X_0$ is the prediction error). We use a similar notation for prediction based on the infinite past. Thus $\mathcal{H}_{(-\infty,-\infty]}$ is the closed linear span (cls) of X_k , $k \leq -1$, $P_{(-\infty,-1]}$ is the corresponding projection, and similarly for other time-intervals. Write

$$\mathcal{H}_n := \mathcal{H}_{(-\infty,n]}$$

for the (subspace generated by) the past up to time n ,

$$\mathcal{H}_{-\infty} := \bigcap_{n=-\infty}^{\infty} \mathcal{H}_n$$

for their intersection, the (subspace generated by) the *remote past*. With $\text{corr}(Y, Z) := E[Y\overline{Z}]/\sqrt{E[|Y|^2].E[|Z|^2]}$ for Y, Z zero-mean and not a.s. 0, write also

$$\alpha_n := \text{corr}(X_n - P_{[1,n-1]}X_n, X_0 - P_{[1,n-1]}X_0)$$

for the correlation between the residuals at times 0, n resulting from (linear) regression on the intermediate values X_1, \dots, X_{n-1} . The sequence

$$\alpha = (\alpha_n)_{n=1}^{\infty}$$

is called the *partial autocorrelation function* (PACF). It is also called the sequence of *Verblunsky coefficients*, for reasons which will emerge below.

Theorem 2 (Verblunsky's Theorem. There is a bijection between the sequences $\alpha = (\alpha_n)$ with each $\alpha_n \in D$ and the probability measures μ on T .

This result dates from Verblunsky in 1936 [V2], in connection with OPUC. It was re-discovered long afterwards by Barndorff-Nielsen and Schou [BarN-S] in 1973 and Ramsey [Ram] in 1974, both in connection with parametrization of time-series models in statistics. The Verblunsky bijection has the great advantage to statisticians of giving an *unrestricted parametrization*: the only restrictions on the α_n are the obvious ones resulting from their being correlations – $|\alpha_n| \leq 1$, or as μ is non-trivial, $|\alpha_n| < 1$. By contrast, $\gamma = (\gamma_n)$ gives a *restricted parametrization*, in that the possible values of γ_n are restricted by the inequalities of positive-definiteness (principal minors of the Toeplitz matrix Γ are positive). This partly motivates the detailed study of the PACF in, e.g., [In1], [In2], [In3], [InKa1], [InKa2]. For general statistical background on partial autocorrelation, see e.g. [KenSt], Ch. 27 (Vol. 2), §46.26-28 (Vol. 3).

As we mentioned in §1, the basic result for OPUC corresponding to Favard's theorem for OPRL is *Szegő's recurrence* (or *recursion*): given a probability measure μ on T , let Φ_n be the monic orthogonal polynomials they generate (by Gram-Schmidt orthogonalization). For every polynomial Q_n of degree n , write

$$Q_n^*(z) := z^n \overline{Q_n(1/\bar{z})}$$

for the *reversed polynomial*. Then the Szegő recursion is

$$\Phi_{n+1}(z) = z\Phi_n(z) - \bar{\alpha}_{n+1}\Phi_n^*(z),$$

where the parameters α_n lie in D :

$$|\alpha_n| < 1,$$

and are the *Verblunsky coefficients* (also known variously as the Szegő, Schur, Geronimus and reflection coefficients; see [Si4], §1.1). The double use of the name Verblunsky coefficients and the notation $\alpha = (\alpha_n)$ for the PACF and the coefficients is justified: the two coincide. Indeed, the Szegő recursion is known in the time-series literature as the *Durbin-Levinson algorithm*; see e.g. [BroDav], §§3.4, 5.2. The term Verblunsky coefficient is from Simon [Si4], to which we refer repeatedly. We stress that Simon writes α_n for our α_{n+1} , and so has $n = 0, 1, \dots$ where we have $n = 1, 2, \dots$. Our notational convention is already established in the time-series literature (see e.g. [BroDav], §§3.4, 5.2), and is more convenient in our context of the PACF, where $n = 1, 2, \dots$ has the direct interpretation as a time-lag between past and future (cf. [Si4], (1.5.15),

p. 56-57). See [Si4], §1.5 and (for two proofs of Verblunsky's theorem) §1.7, 3.1, and [McLZ] for a recent application of the unrestricted PACF parametrization.

One may partially summarize the distributional aspects of Theorems 1 and 2 by the one-one correspondences

$$\alpha \leftrightarrow \mu \leftrightarrow \gamma.$$

The Durbin-Levinson algorithm

Write

$$\hat{X}_{n+1} := \phi_{n1}X_n + \dots + \phi_{nn}X_1$$

for the best linear predictor of X_{n+1} given X_n, \dots, X_1 ,

$$v_n := E[(X_{n+1} - \hat{X}_{n+1})^2] = E[(X_{n+1} - P_{[1,n]}X_{n+1})^2]$$

for the mean-square error in the prediction of X_{n+1} based on X_1, \dots, X_n ,

$$\phi_n := (\phi_{n1}, \dots, \phi_{nn})^T \quad (fpc)$$

for the vector of *finite-predictor coefficients*. The Durbin-Levinson algorithm ([Lev], [Dur]; [BroDav] §5.2, [Pou] §7.2) gives the ϕ_{n+1} , v_{n+1} *recursively*, in terms of quantities known at time n , as follows:

(i) The first component of ϕ_{n+1} is given by

$$\phi_{n+1,n+1} = [\gamma_{n+1} - \sum_{j=1}^n \phi_{nj}\gamma_{n-j}]/v_n.$$

The ϕ_{nn} are the Verblunsky coefficients α_n :

$$\phi_{nn} = \alpha_n.$$

(ii) The remaining components are given by

$$\begin{pmatrix} \phi_{n+1,1} \\ \vdots \\ \phi_{n+1,n} \end{pmatrix} = \begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{pmatrix} - \phi_{n+1,n+1} \begin{pmatrix} \phi_{nn} \\ \vdots \\ \phi_{n1} \end{pmatrix} = \begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{pmatrix} - \alpha_{n+1} \begin{pmatrix} \phi_{nn} \\ \vdots \\ \phi_{n1} \end{pmatrix}.$$

(iii) The prediction errors are given recursively by

$$v_0 = 1, \quad v_{n+1} = v_n[1 - |\phi_{n+1,n+1}|^2] = v_n[1 - |\alpha_{n+1}|^2].$$

In particular, $v_n > 0$ and we have from (ii) that

$$\phi_{nj} - \phi_{n+1,j} = \alpha_{n+1}\phi_{n,n+1-j}. \quad (DL).$$

Since by (iii)

$$v_n = \prod_{j=1}^n [1 - |\alpha_n|^2],$$

the n -step prediction error variance $v_n \rightarrow \sigma^2 > 0$ iff the infinite product converges, that is, $\alpha \in \ell_2$, an important condition that we will meet in §3 below in connection with Szegő's condition.

Note. 1. The Durbin-Levinson algorithm is related to the Yule-Walker equations of time-series analysis (see e.g. [BroDav], §8.1), but avoids the need there for matrix inversion.

2. The computational complexity of the Durbin-Levinson algorithm grows quadratically, rather than cubically as one might expect; see e.g. Golub and van Loan [GolvL], §4.7. Its good numerical properties result from efficient use of the Toeplitz character of the matrix Γ (or equivalently, of Szegő recursion).

3. See [KatSeTe] for a recent approach to the Durbin-Levinson algorithm, and [Deg] for the multivariate case.

Stochastic versus non-stochastic

This paper studies prediction theory for stationary stochastic processes. As an extreme example (in which no prediction is possible), take the ‘free’ case, in which the X_n are independent (and identically distributed). Then $d\mu(\theta) = d\theta/2\pi$, $\gamma_n = \delta_{n0}$, $\alpha_n \equiv 0$, $\Phi_n(z) = z^n$ ([Si4], Ex. 1.6.1).

In contrast to this is the situation where $X = (X_n)$ is *non-stochastic* – deterministic, but (typically) chaotic. This case often arises in non-linear time-series analysis and dynamical systems; for a monograph treatment, see Kantz and Schreiber [KanSch].

One natural way to classify results on OPUC is by the strength of the conditions that they impose. Simon’s book discusses a range of conditions, starting with a fairly weak one, Szegő’s condition ([Si4] Ch. 2 and §3 below), and proceeding to two principal stronger ones, Baxter’s condition ([Si4] Ch. 5 and §4 below) and the strong Szegő condition ([Si4] Ch. 6 and §5 below). From a probabilistic viewpoint, equally important are a range of intermediate conditions not discussed in Simon’s book. These we discuss in §6. We

close with some remarks in §7.

§3. Weak conditions: Szegő's theorem.

Rakhmanov's Theorem

One naturally expects that the influence of the distant past decays with increasing lapse of time. So one wants to know when

$$\alpha_n \rightarrow 0 \quad (n \rightarrow \infty).$$

By *Rakhmanov's theorem* ([Rak]; [Si5] Ch 9, and Notes to §9.1, [MatNeTo]), this happens if the density w of the absolutely continuous component μ_a is positive on a set of full measure:

$$|\{\theta : w(\theta) > 0\}| = 1$$

(using normalized Lebesgue measure – or 2π using Lebesgue measure).

Non-determinism and the Wold decomposition.

Write σ^2 for the one-step mean-square prediction error:

$$\sigma^2 := E[(X_0 - P_{(-\infty, -1]}X_0)^2];$$

by stationarity, this is the $\sigma^2 = \lim_{n \rightarrow \infty} v_n$ above. Call X *non-deterministic* (*ND*) if $\sigma > 0$, *deterministic* if $\sigma = 0$. (This usage is suggested by the usual one of non-randomness being zero-variance, though here a non-deterministic process may be random, but independent of time, so the stochastic process reduces to a random variable.) The *Wold decomposition* (von Neumann [vN] in 1929, Wold [Wo] in 1938; see e.g. Doob [Do], XII.4, Hannan [Ha1], Ch. III) expresses a process X as the sum of a non-deterministic process U and a deterministic process V :

$$X_n = U_n + V_n;$$

the process U is a moving average,

$$U_n = \sum_{j=-\infty}^n m_{n-j} \xi_j = \sum_{k=0}^{\infty} m_k \xi_{n-k},$$

with the ξ_j zero-mean and uncorrelated, with each other and with V ; $E[\xi_n] = 0$, $\text{var}(\xi_n) = E[\xi_n^2] = \sigma^2$. Thus when $\sigma = 0$ the ξ_n are 0, U is missing and the

process is deterministic. When $\sigma > 0$, the spectral measures of U_n, V_n are μ_{ac} and μ_s , the absolutely continuous and singular components of μ . Think of ξ_n as the ‘innovation’ at time n – the new random input, a measure of the unpredictability of the present from the past. This is only present when $\sigma > 0$; when $\sigma = 0$, the present is determined by the past – even by the remote past.

The Wold decomposition arises in operator theory ([vN]; Sz.-Nagy and Foias in 1970 [SzNF], Rosenblum and Rovnyak in 1985 [RoRo], §1.3, [Nik2]), as a decomposition into the unitary and completely non-unitary (cnu) parts.

Szegő's Theorem

Theorem 3 (Szegő's Theorem).

(i) $\sigma > 0$ iff $\log w \in L_1$, that is,

$$\int -\log w(\theta) d\theta > -\infty. \quad (Sz)$$

(ii) $\sigma > 0$ iff $\alpha \in \ell_2$.

(iii)

$$\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2),$$

so $\sigma > 0$ iff the product converges, i.e. iff

$$\sum |\alpha_n|^2 < \infty : \quad \alpha \in \ell_2;$$

(iv) σ^2 is the geometric mean $G(\mu)$ of μ :

$$\sigma^2 = \exp\left(\frac{1}{2\pi} \int \log w(\theta) d\theta\right) =: G(\mu) > 0. \quad (K)$$

Proof. Parts (i), (ii) are due to Szegő [Sz2], [Sz3] in 1920-21, with μ absolutely continuous, and to Verblunsky [V2] in 1936 for general μ . See [Si4] Ch. 2, [Si9] Ch. 2. Parts (iii) and (iv) are due to Kolmogorov in 1941 [Kol]. Thus (K) is called *Kolmogorov's formula*. The alternative name for *Szegő's condition* (Sz) is the *non-determinism condition* (ND), above. //

We now restrict attention to processes for which Szegő's condition holds; indeed, we shall move below to stronger conditions.

The original motivation of Szegő, and later Verblunsky, was *approximation theory*, specifically approximation by polynomials. The Kolmogorov

Isomorphism Theorem allows us to pass between finite sections of the past to polynomials; denseness of polynomials allows prediction with zero error (a ‘bad’ situation – determinism), which happens iff (Sz) does not hold. There is a detailed account of the (rather involved) history here in [Si4] §2.3. Other classic contributions include work of Krein in 1945, Levinson in 1947 [Lev] and Wiener in 1949 [Wi2]. See [BroDav] §5.8 (where un-normalized Lebesgue measure is used, so there is an extra factor of 2π on the right of (K)), [Roz] §II.5 from the point of view of time series, [Si4] for OPUC.

Pure non-determinism, (PND)

When the remote past is trivial,

$$\mathcal{H}_{-\infty} := \bigcap_{n=-\infty}^{\infty} \mathcal{H}_n = \{0\}, \quad (PND)$$

there is no deterministic component in the Wold decomposition, and no singular component in the spectral measure. The process is then called *purely non-deterministic*. Thus

$$(PND) = (ND) + (\mu_s = 0) = (Sz) + (\mu_s = 0) = (\sigma > 0) + (\mu_s = 0) \quad (PND)$$

(usage differs here: the term ‘regular’ is used for (PND) in [IbRo], IV.1, but for (ND) in [Do], XII.2).

The Szegő function and Hardy spaces

Szegő’s theorem is the key result in the whole area, and to explore it further we need the Szegő function (h , below). For this, we need the language and viewpoint of the theory of Hardy spaces, and some of its standard results; several good textbook accounts are cited in §1. For $0 < p < \infty$, the *Hardy space* H_p is the class of analytic functions f on D for which

$$\sup_{r < 1} \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right)^{1/p} < \infty. \quad (H_p)$$

As well as in time series and prediction, as here, Hardy spaces are crucial for martingale theory (see e.g. [Bin1] and the references there). For an entertaining insight into Hardy spaces in probability, see Diaconis [Dia].

For non-deterministic processes, define the *Szegő function* h by

$$h(z) := \exp \left(\frac{1}{4\pi} \int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z} \right) \log w(\theta) d\theta \right) \quad (z \in D), \quad (OF)$$

(note that in [In1-3], [InKa1,2], [Roz] II.5 an extra factor $\sqrt{2\pi}$ is used on the right), or equivalently

$$H(z) := h^2(z) = \exp\left(\frac{1}{2\pi} \int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z}\right) \log w(\theta) d\theta\right) \quad (z \in D).$$

Because $\log w \in L_1$ by (Sz) , H is an *outer function* for H_1 (whence the name *(OF)* above); see Duren [Du], §2.4. By Beurling's canonical factorization theorem,

- (i) $H \in H_1$, the Hardy space of order 1 ([Du], §2.4), or as $H = h^2$, $h \in H_2$.
- (ii) The radial limit

$$H(e^{i\theta}) := \lim_{r \uparrow 1} H(re^{i\theta})$$

exists a.e., and

$$|H(e^{i\theta})| = |h(e^{i\theta})|^2 = w(\theta)$$

(thus h may be regarded as an ‘analytic square root’ of w). See also Hoffman [Ho], Ch. 3-5, Rudin [Ru], Ch. 17, Helson [He], Ch. 4.

Kolmogorov's formula now reads

$$\sigma^2 = m_0^2 = h(0)^2 = G(\mu) = \exp\left(\frac{1}{2\pi} \int \log w(\theta) d\theta\right). \quad (K)$$

When $\sigma > 0$, the Maclaurin coefficients $m = (m_n)$ of the Szegő function $h(z)$ are the moving-average coefficients of the Wold decomposition (recall that the moving-average component does not appear when $\sigma = 0$); see Inoue [In3] and below. When $\sigma > 0$, $m \in \ell_2$ is equivalent to convergence in mean square of the moving-average sum $\sum_{j=0}^{\infty} m_{n-j} \xi_j$ in the Wold decomposition. This is standard theory for orthogonal expansions; see e.g. [Do], IV.4. Note that a function being in H_2 and its Maclaurin coefficients being in ℓ_2 are equivalent by general Hardy-space theory; see e.g. [Ru], 17.10 (see also Th. 17.17 for factorization), [Du] §1.4, 2.4, [Z2], VII.7.

Simon [Si4], §2.8 – ‘Lots of equivalences’ – gives Szegő's theorem in two parts. One ([Si4] Th. 2.7.14) gives twelve equivalences, the other ([Si4], Th. 2.7.15) gives fifteen; the selection of material is motivated by spectral theory [Si5]. Theorem 3 above extends these lists of equivalences, and treats the material from the point of view of probability theory. (It does *not*, however, give a condition on the autocorrelation $\gamma = (\gamma_n)$ equivalent to (Sz) ; this is one of the outstanding problems of the area.)

The contrast here with Verblunsky's theorem is striking. In general, one

has unrestricted parametrization: *all values $|\alpha_n|$ are possible, for all n* . But under Szegő's condition, one has $\alpha \in \ell_2$, and in particular $\alpha_n \rightarrow 0$, as in Rakhmanov's theorem. Thus non-deterministic processes fill out only a tiny part of the α -parameter space D^∞ . One may regard this as showing that the remote past, trivial under (Sz) , has a rich structure in general, as follows:

Szegő's alternative (or dichotomy).

One either has

$$\log w \in L_1 \text{ and } \mathcal{H}_{-\infty} \neq \mathcal{H}_{-n} \neq \mathcal{H}$$

or

$$\log w \notin L_1 \text{ and } \mathcal{H}_{-\infty} = \mathcal{H}_{-n} = \mathcal{H}.$$

In the former case, α occupies a tiny part ℓ_2 of D^∞ , and the remote past $\mathcal{H}_{-\infty}$ is identified with $L_2(\mu_s)$. This is trivial iff $\mu_s = 0$; cf. (PND) . In the second case, α occupies all of D^∞ , and the remote past is the whole space.

Szegő's dichotomy may be interpreted by analogy with physical systems. Some systems (typically, liquids and gases) are 'loose' – left alone, they will thermalize, and tend to an equilibrium in which the details of the past history are forgotten. By contrast, some systems (typically, solids) are 'tight': for example, in tempered steel, the thermal history is locked in permanently by the tempering process. Long memory is also important in economics and econometrics; for background here, see e.g. [Rob], [TeKi].

Note. 1. Our h is the Szegő function D of Simon [Si4], (2.4.2), and $-1/h$ (see below) its negative reciprocal $-\Delta$ [Si4], (2.2.92):

$$h = D, \quad -1/h = -\Delta$$

(we use both notations to facilitate comparison between [In1-3], [InKa1,2], which use h , to within the factor $\sqrt{2\pi}$ mentioned above, and [Si4], our reference on OPUC, which uses D).

2. Both h and $-1/h$ are analytic and non-vanishing in D . See [Si4], Th. 2.2.14 (for $-1/h$, or Δ), Th. 2.4.1 (for h , or D).
3. That (Sz) implies $h = D$ is in the unit ball of H_2 is in [Si4], Th. 2.4.1.
4. See de Branges and Rovnyak [dBR] for general properties of such square-summable power series.
5. Our autocorrelation γ is Simon's c (he calls our γ_n , or his c_n , the *moments* of μ : [Si4], (1.1.20)). Our moving-average coefficients $m = (m_n)$ have no

counterpart in [Si4], and nor do the autoregressive coefficients $r = (r_n)$ or minimality (see below for these). We will also need the Fourier coefficients of $\log w$, known for reasons explained below as the *cepstrum*), which we write as $L = (L_n)$ ('L for logarithm': Simon's \hat{L}_n [Si4], (6.1.13)), and a sequence $b = (b_n)$, the phase coefficients (Fourier coefficients of \bar{h}/h).

6. Lund et al. [LuZhKi] give several properties – monotonicity, convexity etc. – which one of m, γ has iff the other has.

MA(∞) and *AR*(∞)

The power series expansion

$$h(z) = \sum_{n=0}^{\infty} m_n z^n \quad (z \in D)$$

generates the *MA*(∞) coefficients $m = (m_n)$ in the Wold decomposition. That of

$$-1/h(z) = \sum_{n=0}^{\infty} r_n z^n \quad (z \in D)$$

generates the *AR*(∞) coefficients $r = (r_n)$ in the (infinite-order) *autoregression*

$$\sum_{j=-\infty}^n r_{n-j} X_j + \xi_n = 0 \quad (n \in \mathbb{Z}). \quad (AR)$$

See [InKa2] §2, [In3] for background.

One may thus extend the above list of one-one correspondences, as follows:

$$\text{Under } (Sz), \quad \alpha, \mu, \gamma \leftrightarrow m = (m_n) \leftrightarrow h, -1/h \leftrightarrow r = (r_n).$$

Finite and infinite predictor coefficients.

We met the n -vector ϕ_n of finite-predictor coefficients in (*fpc*) of §1; we can extend it to an infinite vector, still denoted ϕ_n , by adding zeros. The corresponding vector $\phi := (\phi_1, \phi_2, \dots)$ of *infinite-predictor coefficients* gives the infinite predictor

$$P_{(-\infty, -1]} X_0 = \sum_{j=1}^{\infty} \phi_j X_{-j} \quad (ipc)$$

([InKa2], (1.4)). One would expect convergence of finite-predictor to infinite-predictor coefficients; under Szegö's condition, one has such convergence in ℓ_2 iff (*PND*), i.e., $\mu_s = 0$:

$$\phi_n \rightarrow \phi \quad \text{in } \ell_2 \Leftrightarrow (PND)$$

(Pourahmadi [Pou], Th. 7.14).

The Szegő limit theorem.

With $G(\mu)$ as above, write T_n (or $T_n(\gamma)$, or $T_n(\mu)$) for the $n \times n$ Toeplitz matrix $\Gamma^{(n)}$ with elements

$$\Gamma_{ij}^{(n)} := c_{j-i}$$

obtained by truncation of the Toeplitz matrix Γ (cf. [BotSi2]). *Szegő's limit theorem* states that, under (Sz) , its determinant satisfies

$$\frac{1}{n} \log \det T_n \rightarrow G(\mu) \quad (n \rightarrow \infty)$$

(note that (Sz) is needed for the right to be defined). A stronger statement – *Szegő's strong limit theorem* – holds; we defer this till §5.

The Szegő limit theorem is used in the *Whittle estimator* of time-series analysis; see e.g. Whittle [Wh], Hannan [Ha2].

Phase coefficients.

When the Szegő condition (Sz) holds, the Szegő function $h(z) = \sum_0^\infty m_n z^n$ is defined. We can then define the *phase function* \bar{h}/h , so called because it has unit modulus and depends only on the phase or argument of h (Peller [Pel], §8.5). Its Fourier coefficients b_n are called the *phase coefficients*. They are given in terms of $m = (m_n)$ and $r = (r_n)$ by

$$b_n := \sum_0^\infty m_k r_{n+k} \quad (n = 0, 1, 2, \dots). \quad (b)$$

The role of the phase coefficients is developed in [BiInKa]. They are important in connection with rigidity (§6 below), and Hankel operators [Pel].

Rajchman measures.

In the Gaussian case, mixing in the sense of ergodic theory holds iff

$$\gamma_n \rightarrow 0 \quad (n \rightarrow \infty)$$

([CorFoSi], §14.2, Th. 2). Since (Sz) is $\gamma \in \ell_2$, which implies $\gamma_n \rightarrow 0$, this is even weaker than (Sz) . Measures for which this condition holds are called *Rajchman measures* (they were studied by A. Rajchman in the 1920s). Here the continuous singular part μ_{cs} of μ is decisive; for a characterization of

Rajchman measures, see Lyons ([Ly1] – [Ly3] and the appendix to [KahSa]).

$ARMA(p, q)$.

The Box-Jenkins $ARMA(p, q)$ methodology ([BoxJeRe], [BroDav]: autoregressive of order p , moving average of order q – see §6.3 for $MA(q)$) applies to stationary time series where the roots of the relevant polynomials lie in the unit disk (see e.g. [BroDav] §3.1). The limiting case, of *unit roots*, involves non-stationarity, and so the statistical dangers of spurious regression (§1); cf. Robinson [Rob], p.2. We shall meet other instances of unit-root phenomena later (§6.3).

Szegö's theorem and the Gibbs Variational Principle

We point out that Verblunsky [V2] proved the Gibbs Variational Principle, one of the cornerstones of nineteenth-century statistical mechanics, for the Szegö integral:

$$\inf_g \left[\int e^g d\mu / \exp \left(\int g d\theta / 2\pi \right) \right] = \exp \left[\int \log w(\theta) d\theta / 2\pi \right].$$

For details, see e.g. Simon [Si9] §§2.2, 10.6, [Si10], Ch. 16, 17. For background on the Gibbs Variational Principle, see e.g. Simon [Si1], III.4, Georgii [Geo], 15.4, Ellis [Ell], III.8.

§4. Strong conditions: Baxter's theorem

The next result ([Bax1], [Bax2], [Bax3]; [Si4], Ch. 5) gives the first of our strong conditions.

Theorem 4 (Baxter's theorem). The following are equivalent:

(i) the Verblunsky coefficients (or PACF) are summable,

$$\alpha \in \ell_1; \tag{B}$$

(ii) the autocorrelations are summable, $\gamma \in \ell_1$, and μ is absolutely continuous with continuous positive density:

$$\min_{\theta} w(\theta) > 0.$$

Of course, (γ_n) summable gives, as the γ_n are the Fourier coefficients of μ , that μ is absolutely continuous with continuous density w ; thus $w > 0$

iff $\inf w = \min w > 0$.) We extend this list of equivalences, and bring out its probabilistic significance, in Theorem 5 below on ℓ_1 (this is substantially Theorem 4.1 of [In3]). We call $\alpha \in \ell_1$ (or any of the other equivalences in Theorem 4) *Baxter's condition* (whence (B) above). Since $\ell_1 \subset \ell_2$, Baxter's condition (B) ('strong') implies Szegő's condition (Sz) ('weak').

Theorem 5 (Inoue). For a stationary process X , the following are equivalent:

- (i) Baxter's condition (B) holds: $\alpha \in \ell_1$.
- (ii) $\gamma \in \ell_1$, $\mu_s = 0$ and the spectral density w is continuous and positive.
- (iii) (PND) (that is, $(Sz)/(ND) + \mu_s = 0$) holds, and the moving-average and autoregressive coefficients are summable:

$$m \in \ell_1, \quad r \in \ell_1.$$

- (iv) $m \in \ell_1$, $\mu_s = 0$ and the spectral density w is continuous and positive.
- (v) $r \in \ell_1$, $\mu_s = 0$ and the spectral density w is continuous and positive.

Proof.

(i) \Leftrightarrow (ii). This is Baxter's theorem, as above.

(iii) \Rightarrow (iv), (v). By (PND) , (Sz) holds, so the non-tangential limit

$$h(e^{i\theta}) = \lim_{r \uparrow 1} h(re^{i\theta}) = \sum_{n=0}^{\infty} m_n e^{in\theta}$$

exists a.e. But as $m \in \ell_1$, $h(e^{i\theta})$ is continuous, so this holds everywhere. Since

$$w(\theta) = |h(e^{i\theta})|^2 = |D(e^{i\theta})|^2 = \left| \sum_{n=0}^{\infty} m_n e^{in\theta} \right|^2,$$

w is continuous. Letting $r \uparrow 1$ in

$$h(z)(-1/h(z)) = \left(\sum_0^{\infty} m_n r^n e^{in\theta} \right) \left(\sum_0^{\infty} r_n r^n e^{in\theta} \right) = -1$$

gives similarly

$$\left(\sum_0^{\infty} m_n e^{in\theta} \right) \left(\sum_0^{\infty} r_n e^{in\theta} \right) = -1.$$

So $h(e^{i\theta})$ has no zeros, so neither does w . That is, (iv), (v) hold.

(iv) \Rightarrow (iii). As w is positive and continuous, w is bounded away from 0 and

∞ . So $1/w$ is also. So

$$1/w(\theta) = |1/h(e^{i\theta})|^2 = |\Delta(e^{i\theta})|^2 = \left| \sum_{n=0}^{\infty} r_n e^{in\theta} \right|^2,$$

where $\Delta = 1/D$. (See [Si4], Th. 2.2.14, 2.7.15: the condition $\lambda_{\infty}(\cdot) > 0$ there is (Sz) , so holds here.) By Wiener's theorem, the reciprocal of a non-vanishing absolutely convergent Fourier series is an absolutely convergent Fourier series (see e.g. [Ru], Th. 18.21). So from $m \in \ell_1$ we obtain $r \in \ell_1$, whence (iii) (cf. [Berk], p.493).

(v) \Rightarrow (iii). This follows as above, by Wiener's theorem again.

(iv) \Rightarrow (ii). From the $MA(\infty)$ representation,

$$\gamma_n = \sum_{k=0}^{\infty} m_{|n|+k} m_k \quad (n \in \mathbb{Z}) \quad (conv)$$

([InKa2], (2.21)). So as ℓ_1 is closed under convolution, $m \in \ell_1$ implies $\gamma \in \ell_1$, indeed with

$$\|\gamma\|_1 \leq \|m\|_1^2,$$

giving (ii).

(ii) \Rightarrow (v). We have

$$\phi_j = c_0 r_j = \sigma r_j$$

with ϕ_j the infinite-predictor coefficients ([InKa2], (3.1)). Then $r \in \ell_1$ follows by the Wiener-Lévy theorem, as in Baxter [Ba3], 139-140. //

Note. 1. Under Baxter's condition, both $|h|$ and $|1/h|$ (or $|D|$ and $|\Delta| = |1/D|$) are continuous and positive on the unit circle. As $h, 1/h$ are analytic in the disk, and so attain their maximum modulus on the circle by the maximum principle,

$$\inf_D |h(\cdot)| > 0, \quad \inf_D |1/h(\cdot)| > 0$$

(and similarly for $D(\cdot), \Delta$); [Si4], (5.2.3), (5.2.4).

2. The hard part of Baxter's theorem is (ii) \Rightarrow (i), as Simon points out ([Si4], 314).

3. Simon [Si4], Th. 5.2.2 gives *twelve* equivalences in his final form of Baxter's theorem. (He does not, however, deal explicitly with m and r .)

4. Simon also gives a more general form, in terms of *Beurling weights*, ν . The relevant Banach algebras contain the Wiener algebra used above as the

special case $\nu = 1$.

5. The approach of [Si4], §5.1 is via truncated Toeplitz matrices and their inverses. The method derives, through Baxter's work, from the Wiener-Hopf technique. This point of view is developed at length in [BotSi1], [BotSi2]. Baxter's motivation was approximation to infinite-past predictors by finite-past predictors.

Long-range dependence

In various physical models, the property of *long-range dependence* (LRD) is important, particularly in connection with phase transitions (see e.g. [Si1], Ch. II, [Gri1], Ch. 9, [Gri2], Ch. 5), to which we return below. This is a spatial property, but applies also in time rather than space, when the term used is *long memory*. A good survey of long-memory processes was given by Cox [Cox] in 1984, and a monograph treatment by Beran [Ber] in 1994. For more recent work, see [DouOpTa], [Rob], [Gao] Ch. 6, [TeKi], [GiKoSu].

Baxter's theorem is relevant to the definition of LRD recently proposed independently by Debowski [Deb] and Inoue [In3]: *long-range dependence, or long memory, is non-summability of the PACF*:

$$X \text{ has LRD iff } \alpha \notin \ell_1. \quad (DI)$$

While the broad concept of long memory, or LRD, has long been widely accepted, authors differed over the precise definition. There were two leading candidates:

- (i) LRD is non-summability of covariances, $\gamma \notin \ell_1$.
- (ii) LRD is covariance decaying like a power: $\gamma_n \sim c/n^{1-2d}$ as $n \rightarrow \infty$, for some parameter $d \in (0, 1/2)$ (d for differencing – see below) and constant $c \in (0, \infty)$ (and so $\sum \gamma_n = \infty$).

Note. 1. In place of (ii), one may require $w(\theta) \sim C/\theta^{2d}$ as $\theta \downarrow 0$, for some constant $C \in (0, \infty)$. The constants here may be replaced by slowly varying functions. See e.g. [BinGT] §4.10 for relations between regular variation of Fourier series and Fourier coefficients.

2. One often encounters, instead of $d \in (0, 1/2)$, a parameter $H = d + \frac{1}{2} \in (1/2, 1)$. This H is the *Hurst parameter*, named after the classic studies by the hydrologist Hurst of water flows in the Nile; see [Ber], Ch. 2.

3. For $d \in (0, \frac{1}{2})$, $\ell(\cdot)$ slowly varying, the following class of prototypical long-memory examples is considered in [InKa2], §2.3 (see also [In1], Th. 5.1):

$$\gamma_n \sim \ell(n)^2 B(d, 1 - 2d) / n^{1-2d},$$

$$m_n \sim \ell(n)/n^{1-d},$$

$$r_n \sim \frac{d \sin(\pi d)}{\pi} \cdot \frac{1}{\ell(n)} \cdot 1/n^{1+d}.$$

See the sources cited for inter-relationships between these.

4. In [InKa2], Example 2.6, the class of $FARIMA(p, d, q)$ processes is considered (obtained from an $ARIMA(p, q)$ process by fractional differencing of order d – see [Hos], [BroDav], [KokTa]). For $d \in (0, 1/2)$ these have long memory; for $d = 0$ they reduce to the familiar $ARMA(p, q)$ processes.

Li ([Li], §3.4) has recently given a related but different definition of long memory; we return to this in §5 below.

5. Strong conditions: the strong Szegő theorem

The work of this section may be motivated by work from two areas of physics.

1. The cepstrum.

During the Cold War, the problem of determining the signature of the underground explosion in a nuclear weapon test, and distinguishing it from that of an earthquake, was very important, and was studied by the American statistician J. W. Tukey and collaborators. Write $L = (L_n)$, where the L_n are the Fourier coefficients of $\log w$, the log spectral density:

$$L_n := \int \log w(\theta) e^{in\theta} d\theta / 2\pi.$$

Thus $\exp(L_0)$ is the geometric mean $G(\mu)$. The sequence L is called the *cepstrum*, L_n the *ceptral coefficients* (Simon's notation here is \hat{L}_n ; [Si4], (2.1.14), (6.1.11)); see e.g. [OpSc], Ch. 12. The terminology dates from work of Bogert, Healy and Tukey of 1963 on echo detection [BogHeTu]; see McCullagh [McC], Brillinger [Bri] (the term is chosen to suggest both echo and spectrum, by reversing the first half of the word spectrum; it is accordingly pronounced with the c hard, like a k).

2. The strong Szegő limit theorem.

This (which gives the weak form on taking logarithms) states (in its present form, due to Ibragimov) that

$$\frac{\det T_n}{G(\mu)^n} \rightarrow E(\mu) := \exp\left\{\sum_{k=1}^{\infty} k L_k^2\right\} \quad (n \rightarrow \infty)$$

(of course the sum here must converge; it turns out that this form is best-possible: the result is valid whenever it makes sense ([Si4], 337).

The motivation was Onsager's work in the two-dimensional Ising model, and in particular *Onsager's formula*, giving the existence of a critical temperature T_c and the decay of the magnetization as the temperature $T \uparrow T_c$; see [BotSi2] §5.1, [Si1] II.6, [McCW]. The mechanism was a question by Onsager (c. 1950) to his Yale colleague Kakutani, who asked Szegö ([Si4], 331).

Write $H^{1/2}$ for the subspace of ℓ_2 of sequences $a = (a_n)$ with

$$\|a\|^2 := \sum_n (1 + |n|) |\alpha_n|^2 < \infty \quad (H^{1/2})$$

(the function of the '1' on the right is to give a norm; without it, $\|\cdot\|$ vanishes on the constant functions). This is a Sobolev space ([Si4], 329, 337; it is also a Besov space, whence the alternative notation $B_2^{1/2}$; see e.g. Peller [Pel], Appendix 2.6 and §7.13). This is the space that plays the role here of ℓ_2 in §2 and ℓ_1 in §3. Note first that, although ℓ_1 and $H^{1/2}$ are close in that a sequence (n^c) of powers belongs to both or neither, neither contains the other (consider $a_n = 1/(n \log n)$, $a_n = 1/\sqrt{n}$ if $n = 2^k$, 0 otherwise).

Theorem 6 (Strong Szegö Theorem).

(i) If (PND) holds (i.e. $(Sz) = (ND)$ holds and $\mu_s = 0$), then

$$E(\mu) = \prod_{j=1}^{\infty} (1 - |\alpha_j|^2)^{-j} = \exp\left(\sum_{n=1}^{\infty} n L_n^2\right)$$

(all three may be infinite), with the infinite product converging iff the *strong Szegö condition*

$$\alpha \in H^{1/2}, \quad (sSz)$$

holds.

(ii) (sSz) holds iff

$$L \in H^{1/2} \quad (sSz')$$

holds.

(iii) Under (Sz) , finiteness of any (all three) of the expressions in (i) forces $\mu_s = 0$.

Proof. Part (i) is due to Ibragimov ([Si4], Th. 6.1.1), and (ii) is immediate from this. Part (iii) is due to Golinski and Ibragimov ([Si4], Th. 6.1.2; cf.

[Si2]). //

Part of Ibragimov's theorem was recently obtained independently by Li [Li], under the term *reflectrum identity* (so called because it links the Verblunsky or reflection coefficients with the cepstrum), based on information theory – mutual information between past and future. Earlier, Li and Xie [LiXi] had shown the following:

- (i) a process with given autocorrelations $\gamma_0, \dots, \gamma_p$ with minimal information between past and future must be an autoregressive model $AR(p)$ of order p ;
- (ii) a process with given cepstral coefficients L_0, \dots, L_p with minimal information between past and future must be a *Bloomfield model* $BL(p)$ of order p ([Bl1], [Bl2]), that is, one with spectral density $w(\theta) = \exp\{L_0 + 2 \sum_{k=1}^p L_k \cos k\theta\}$.

Another approach to the strong Szegő limit theorem, due to Kac [Kac], uses the conditions

$$\inf w(\cdot) > 0, \quad \gamma = (\gamma_n) \in \ell_1, \quad \gamma \in H^{1/2}$$

(recall that ℓ_1 and $H^{1/2}$ are not comparable). This proof, from 1954, is linked to probability theory – Spitzer's identity of 1956, and hence to fluctuation theory for random walks, for which see e.g. [Ch], Ch. 8.

The Borodin-Okounkov formula.

This turns the strong Szegő limit theorem above from analysis to algebra by identifying the quotient on the left there as a determinant which visibly tends to 1 as $n \rightarrow \infty$ [BorOk]; see [Si4] §6.2. (It was published in 2000, having been previously obtained by Geronimo and Case [GerCa] in 1979; see [Si4] 337, 344, [Bot] for background here.) In terms of operator theory and in Widom's notation [Bot], the result is

$$\frac{\det T_n(a)}{G(a)^n} = \frac{\det(I - Q_n H(b) H(\tilde{c}) Q_n)}{\det(I - H(b) H(\tilde{c}))},$$

for a a sufficiently smooth function without zeros on the unit circle and with winding number 0. Then a has a Wiener-Hopf factorization $a = a_- a_+$; $b := a_- a_+^{-1}$, $c := a_-^{-1} a_+$; $H(b)$, $H(\tilde{c})$ are the Hankel matrices $H(b) = (b_{j+k+1})_{j,k=0}^\infty$, $H(\tilde{c}) = (c_{-j-k-1})_{j,k=0}^\infty$, and Q_n is the orthogonal projection of $\ell^2(1, 2, \dots)$ onto $\ell^2(\{n, n+1, \dots\})$. By Widom's formula,

$$1/\det(I - H(b) H(\tilde{c})) = \exp\left\{\sum_{k=1}^\infty k L_k^2\right\} =: E(a)$$

(see e.g. [Si4], Th. 6.2.13), and $Q_n H(b) H(\tilde{c}) Q_n \rightarrow 0$ in the trace norm, whence

$$\det T_n(a)/G(a)^n \rightarrow E(a),$$

the strong Szegö limit theorem. See [Si4], Ch. 6, [Si6], [BasW], [BotW] (in [Si4] §6.2 the result is given in OPUC terms; here b, c are the phase function \bar{h}/h and its inverse).

$(B + sSz)$.

We may have both of the strong conditions (B) and (sSz) (as happens in Kac's method [Kac], for instance). Matters then simplify, since the spectral density w is now continuous and positive. So w is bounded away from 0 and ∞ , so $\log w$ is bounded. Write

$$\omega^2(\delta, h) := \sup_{|\theta| \leq \delta} \left(\int |h(\lambda + \theta) - h(\lambda)|^2 d\lambda \right)^{1/2}$$

for the L_2 modulus of continuity. Applying [IbRo], IV.4, Lemma 7 to $\log w$,

$$L \in H^{1/2} \Leftrightarrow \sum_{k=1}^{\infty} \omega^2(1/k, \log w) < \infty,$$

and applying it to w ,

$$\gamma \in H^{1/2} \Leftrightarrow \sum_{k=1}^{\infty} \omega^2(1/k, w) < \infty.$$

Thus under (B) , $L \in H^{1/2}$ and $\gamma \in H^{1/2}$ become equivalent. This last condition is Li's proposed definition of long-range dependence:

$$LRD \Leftrightarrow \gamma \notin H^{1/2} \quad (Li)$$

([Li], §3.4; compare the Debowski-Inoue definition (DI) above, that LRD iff $\alpha \notin \ell_1$).

We are now in $W \cap H^{1/2}$, the intersection of $H^{1/2}$ with the Wiener algebra W (of absolutely convergent Fourier series) relevant to Baxter's theorem as in §3. As there, we can take inverses, since the Szegö function is non-zero on the circle (cf. [BotSi2], §5.1). One can thus extend Theorem 2 to this situation, including the cepstral condition $L \in H^{1/2}$ (Li [Li], Th. 1 part 3, showed that $L \in H^{1/2}$ and $\gamma \in H^{1/2}$ are equivalent if w is continuous and

positive).

$L_\infty + (sSz)$.

The bounded functions in $H^{1/2}$ form an algebra, the *Krein algebra* K , a Banach algebra under convolution; see Krein [Kr], Böttcher and Silbermann [BotSi1] Ch. 10, [BotSi2] Ch. 5, [Si4], 344, [BotKaSi]. The Krein algebra may be used as a partial substitute for the Wiener algebra W used to treat Baxter's theorem in §3 ($W \cap H^{1/2}$ is also an algebra: [BotSi], §5.1).

5.1. ϕ -mixing

Weak dependence may be studied by a hierarchy of mixing conditions; for background, see e.g. Bradley [Bra1], [Bra2], [Bra3], Bloomfield [Bl3], Ibragimov and Linnik [IbLi], Ch. 17, Cornfeld et al. [CorFoSi]), and in the Gaussian case Ibragimov and Rozanov [IbRo], Peller [Pel]. We need two sequences of mixing coefficients:

$$\phi(n) := E \sup\{|P(A|\mathcal{F}_{-\infty}^0) - P(A)| : A \in \mathcal{F}_n^\infty\};$$

$$\rho(n) := \rho(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty),$$

where

$$\rho(\mathcal{A}, \mathcal{B}) := \sup\{\|E(f|\mathcal{B}) - Ef\|_2 / \|f\|_2 : f \in L_2(\mathcal{A})\}.$$

The process is called ϕ -mixing if $\phi(n) \rightarrow 0$ as $n \rightarrow \infty$, ρ -mixing if $\rho(n) \rightarrow 0$. (The reader is warned that some authors use other letters here – e.g. [IbRo] uses β for our ϕ ; we follow Bradley.)

We quote [Bra1] that ϕ -mixing implies ρ -mixing. We regard the first as a strong condition, so include it here, but the second and its several weaker relatives as intermediate conditions, which we deal with in §6 below.

The spectral characterization for ϕ -mixing is

$$\mu_s = 0, \quad w(\theta) = |P(e^{i\theta})|^2 w^*(\theta),$$

where P is a polynomial with its roots on the unit circle and the cepstrum $L^* = (L_n^*)$ of w^* satisfies the strong Szegő condition (sSz) ([IbRo] IV.4, p. 129). This is weaker than (sSz). In the Gaussian case, ϕ -mixing (also known as absolute regularity) can also be characterized in operator-theoretic terms: $\phi(n)$ can be identified as $\sqrt{\text{tr}(B_n)}$, where B_n are compact operators with finite trace, so ϕ -mixing is $\text{tr}(B_n) \rightarrow 0$ ([IbRo], IV.2 Th. 4, IV.3 Th. 6).

6. Intermediate conditions

We turn now to four intermediate conditions, in decreasing order of strength.

6.1. ρ -mixing

The spectral characterization of ρ -mixing (also known as complete regularity) is

$$\mu_s = 0, \quad w(\theta) = |P(e^{i\theta})|^2 w^*(\theta),$$

where P is a polynomial with its roots on the unit circle and

$$\log w^* = u + \tilde{v},$$

with u, v real and continuous (Sarason [Sa2]; Helson and Sarason [HeSa]). An alternative spectral characterization is

$$\mu_s = 0, \quad w(\theta) = |P(e^{i\theta})|^2 w^*(\theta),$$

where P is a polynomial with its roots on the unit circle and for all $\epsilon > 0$,

$$\log w^* = r_\epsilon + u_\epsilon + \tilde{v}_\epsilon,$$

where r_ϵ is continuous, u_ϵ, v_ϵ are real and bounded, and $\|u_\epsilon\| + \|v_\epsilon\| < \epsilon$ ([IbRo], V.2 Th. 3; we note here that inserting such a polynomial factor preserves complete regularity, merely changing ρ – [IbRo] V.1, Th. 1).

6.2. Positive angle: the Helson-Szegö and Helson-Sarason conditions.

We turn now to a weaker condition. For subspaces A, B of \mathcal{H} , the *angle* between A and B is defined as

$$\cos^{-1} \sup\{|(a, b)| : a \in A, b \in B\}.$$

Then A, B are at a *positive angle* iff this supremum is < 1 . One says that the process X satisfies the *positive angle condition*, (PA) , if for some time lapse k the past $cls(X_m : m < 0)$ and the future $cls(X_{k+m} : m \geq 0)$ are at a positive angle, i.e. $\rho(0) = \dots \rho(k-1) = 1, \rho(k) < 1$, which we write as $PA(k)$ (Helson and Szegö [HeSz], $k = 1$; Helson and Sarason [HeSa], $k > 1$). The spectral characterization of this is

$$\mu_s = 0, \quad w(\theta) = |P(e^{i\theta})|^2 w^*(\theta),$$

where P is a polynomial of degree $k - 1$ with its roots on the unit circle and

$$\log w^* = u + \tilde{v},$$

where u, v are real and bounded and $\|v\| < \pi/2$ ([IbRo] V.2, Th. 3, Th. 4). (The role of $\pi/2$ here stems from Zygmund's theorem of 1929, that if u is bounded and $\|u\| < \pi/2$, $\exp\{\tilde{u}\} \in L_1$ ([Z1], [Z2] VII, (2.11), [Tor], V.3: cf. [Pel] §3.2.) Thus ρ -mixing implies (PA) (i.e. $PA(k)$ for some k).

The case $PA(k)$ for $k > 1$ is a unit-root phenomenon (cf. the note at the end of §3). We may (with some loss of information) reduce to the case $PA(1)$ by sampling only at every k th time point (cf. [Pel], §§8.5, 12.8). We shall do this for convenience in what follows.

It turns out that the Helson-Szegö condition $(PA(1))$ coincides with *Muckenhoupt's condition* A_2 in analysis:

$$\sup_I \left(\frac{1}{|I|} \int_I w(\theta) d\theta \right) \left(\frac{1}{|I|} \int_I \frac{1}{w(\theta)} d\theta \right) < \infty, \quad (A_2)$$

where $|\cdot|$ is Lebesgue measure and the supremum is taken over all subintervals I of the unit circle T . See e.g. Hunt, Muckenhoupt and Wheeden [HuMuWh]. With the above reduction of PA to $PA(1)$, we then have ρ -mixing implies $PA(1)$ ($= A_2$).

6.3. Pure minimality

Consider now the *interpolation problem*, of finding the best linear interpolation of a missing value, X_0 say, from the others. Write

$$H'_n := \text{cls}\{X_m : m \neq n\}$$

for the closed linear span of the values at times other than n . Call X *minimal* if

$$X_n \notin H'_n,$$

purely minimal if

$$\bigcap_n H'_n = \{0\}.$$

The spectral condition for minimality is (Kolmogorov in 1941, [Kol] §10)

$$1/w \in L_1, \quad (\min)$$

and for pure minimality, $\mu_s = 0$ also (Makagon-Weron in 1976, [MakWe]; Sarason in 1978, [Sa1]; [Pou], Th. 8.10):

$$1/w \in L_1, \quad \mu_s = 0. \quad (\text{purmin})$$

Of course (A_2) implies $1/w \in L_1$, so the Helson-Szegö condition $(PA(1))$ (or Muckenhoupt condition (A_2)) implies pure minimality. (From $\log x < x - 1$ for $x > 1$, (min) implies (Sz) : both restrict the small values of $w \geq 0$, and in particular force $w > 0$ a.e.) For background on the implication from the Helson-Szegö condition $PA(1)$ to (A_2) , see e.g. Garnett [Gar], Notes to Ch. VI, Treil and Volberg [TrVo2].

Under minimality, the relationship between the moving-average coefficients $m = (m_n)$ and the autoregressive coefficients $r = (r_n)$ becomes symmetrical, and one has the following complement to Theorem 4:

Theorem 7 (Inoue). For a stationary process X , the following are equivalent:

- (i) The process is minimal.
- (ii) The autoregressive coefficients $r = (r_n)$ in (AR) satisfy $r \in \ell_2$.
- (iii) $1/h \in H_2$.

Proof. Since

$$1/h(z) = \exp\left(\frac{1}{4\pi} \int \left(\frac{e^{i\theta} + z}{e^{i\theta} - z}\right) \log(1/w(\theta)) d\theta\right) \quad (z \in D), \quad (OF')$$

and $\pm \log w$ are in L_1 together, when $1/w \in L_1$ (i.e. the process is minimal) one can handle $1/w$, $1/h$, $m = (m_n)$ as we handled w , h and $r = (r_n)$, giving

$$1/h \in H_2$$

and

$$r = (r_n) \in \ell_2.$$

Conversely, each of these is equivalent to (min) ; [In1], Prop. 4.2. //

6.4. Rigidity; (LM) , (CND) , (IPF) .

Rigidity; the Levinson-McKean condition.

Call $g \in H^1$ *rigid* if is determined by its phase or argument:

$$f \in H^1 \quad (f \text{ not identically } 0), \quad f/|f| = g/|g| \quad a.e. \quad \Rightarrow$$

$f = cg$ for some positive constant c .

This terminology is due to Sarason [Sa1], [Sa2]; the alternative terminology, due to Nakazi, is *strongly outer* [Na1], [Na2]. One could instead say that such a function is *determined by its phase*. The idea originates with de Leeuw and Rudin [dLR] and Levinson and McKean [LevMcK]. In view of this, we call the condition that μ be absolutely continuous with spectral density $w = |h|^2$ with h^2 rigid, or determined by its phase, the *Levinson-McKean condition*, *(LM)*.

Complete non-determinism; intersection of past and future.

In [InKa2], the following two conditions are discussed:

(i) *complete non-determinism*,

$$\mathcal{H}_{(-\infty, -1]} \cap \mathcal{H}_{[0, \infty)} = \{0\} \quad (CND)$$

(for background on this, see [BlJeHa], [JeBl], [JeBlBa]),

(ii) the *intersection of past and future* property,

$$\mathcal{H}_{(-\infty, -1]} \cap \mathcal{H}_{[-n, \infty)} = \mathcal{H}_{[-n, -1]} \quad (n = 1, 2, \dots) \quad (IPF)$$

These are shown to be equivalent in [InKa2]. In [KaBi], it is shown that both are equivalent to the Levinson-McKean condition, or rigidity:

$$(LM) \quad \Leftrightarrow \quad (IPF) \quad \Leftrightarrow \quad (CND).$$

These are weaker than pure minimality ([Bl3], §7, [KaBi]). But since *(CND)* was already known to be equivalent to *(PND)* + *(IPF)*, they are stronger than *(PND)*. This takes us from the weakest of the four intermediate conditions of this section to the stronger of the weak conditions of §3.

7. Remarks

1. $VMO \subset BMO$.

The spectral characterizations given above were mainly obtained before the work of Fefferman [Fe] in 1971, Fefferman and Stein [FeSt] in 1972 (see Garnett [Gar], Ch. VI for a textbook account): in particular, they predate the *Fefferman-Stein decomposition* of a function of bounded mean oscillation, $f \in BMO$, as

$$f = u + \tilde{v}, \quad u, v \in L_\infty.$$

This has a complement due to Sarason [Sa3], where f here is in VMO iff u, v are continuous. Sarason also gives ([Sa3], Th. 2) a characterization of his class of functions of vanishing mean oscillation VMO within BMO related to Muckenhoupt's condition (A_2) .

While both components u, v are needed here, and may be large in norm, it is important to note that the burden of being large in norm may be born by a *continuous* function, leaving u and \tilde{v} together to be small in (L_∞) norm (in particular, less than $\pi/2$). This is the Ibragimov-Rozanov result ([IbRo], V.2 Th. 3), used in §6.1 to show that absolute regularity (§5) implies complete regularity.

2. $H^{1/2} \subset VMO$.

The class $H^{1/2}$ is contained densely within VMO (Prop. A2, Boutet de Monvel-Berthier et al. [BouGePu]). For $H^{1/2}$, one has a version of the Fefferman-Stein decomposition for BMO:

$$f \in H^{1/2} \quad \Leftrightarrow \quad f = u + \tilde{v}, \quad u, v \in H^{1/2} \cap L^\infty$$

([Pel] §7.13).

3. *Winding number and index.*

The class $H^{1/2}$ occurs in recent work on *topological degree* and winding number; see Brezis [Bre], Bourgain and Kozma [BouKo]. The winding number also occurs in operator theory as an index in applications of Banach-algebra methods and the Gelfand transform; see e.g. [Si4], Ch. 5 (cf. Tsirelson [Ts]).

4. *Conformal mapping.*

The class $H^{1/2}$ also occurs in work of Zygmund on conformal mapping ([Z2], VII.10).

5. *Rapid decay and continuability.*

Even stronger than the strong conditions considered here in §§4, 5 is assuming that the Verblunsky coefficients are rapidly decreasing. This is connected to analytic continuability of the Szegő function beyond the unit disk; see [Si7].

6. *Scattering theory.*

The implication from the strong Szegő (or Golinskii-Ibragimov) condition

to the Helson-Szegö/Helson-Sarason condition (PA) has a recent analogue in scattering theory (Golinskii et al. [GolKhPeYu], under ' GI implies (HS) ').

7. *Wavelets.*

Traditionally, the subject of time series seemed to consist of two non-intercommunicating parts, 'time domain' and 'frequency domain' (known to be equivalent to each other via the Kolmogorov Isomorphism Theorem of §2). The subject seemed to suffer from schizophrenia (see e.g. [BriKri] and [HaKR]) – though the constant relevance of the spectral or frequency side to questions involving time directly is well illustrated in the apt title 'Past and future' of the paper by Helson and Sarason [HeSa] (cf. [Pel] §8.6). This unfortunate schism has been healed by the introduction of *wavelet* methods (see e.g. the standard work Meyer [Me], Meyer and Coifman [MeCo], and in OPUC, Treil and Volberg [TrVo1]). The practical importance of this may be seen in the digitization of the FBI's finger-print data-bank (without which the US criminal justice system would long ago have collapsed). Dealing with time and frequency together is also crucial in other areas, e.g. in the high-quality reproduction of classical music.

8. *Higher dimensions: matrix OPUC (MOPUC).*

We present the theory here in one dimension for simplicity, reserving the case of higher dimensions for a sequel [Bin2]. We note here that in higher dimensions the measure μ and the Verblunsky coefficients α_n become matrix-valued (matrix OPUC, or MOPUC), so one loses commutativity. The multidimensional case is needed for *portfolio theory* in mathematical finance, where one holds a (preferably balanced) portfolio of risky assets rather than one; see e.g. [BinFrKi].

9. *Non-commutativity.*

Much of the theory presented here has a non-commutative analogue in operator theory; see Blecher and Labuschagne [BiLa], Bekjan and Xu [BeXu] and the references cited there.

10. *Non-stationarity.*

As mentioned in §1, the question of whether or not the process is stationary is vitally important, and stationarity is a strong assumption. The basic Kolmogorov Isomorphism Theorem can be extended beyond the stationary case in various ways, e.g. to harmonisable processes (see e.g. [Rao]). For

background, and applications to filtering theory, see e.g. [Kak]; for filtering theory, we refer to e.g. [BaiCr].

11. *Continuous time.*

The Szegő condition (Sz) for the unit circle (regarded as the boundary of the unit disc) corresponds to the condition

$$\int_{-\infty}^{\infty} \frac{\log |f(x)|}{1+x^2} dx > -\infty$$

for the real line (regarded as the boundary of the upper half-plane). This follows from the Möbius function $w = (z-i)/(z+i)$ mapping the half-plane conformally onto the disc; see e.g. [Du], 189-190. The consequences of this condition are explored at length in Koosis' monograph on the 'logarithmic integral', [Koo2]. Passing from the disc to the half-plane corresponds probabilistically to passing from discrete to continuous time (and analytically to passing from Fourier series to Fourier integrals). The probabilistic theory is considered at length in Dym and McKean [DymMcK].

12. *Gaussianity and linearity.*

We have mentioned the close links between Gaussianity and linearity in §1. For background on Gaussian Hilbert spaces and Fock space, see Janson [Jan], Peller [Pel]; for extensions to §§5.1, 6 in the Gaussian case, see [IbRo], [Pel], [Bra1] §5. To return to the undergraduate level of our opening paragraph: for an account of Gaussianity, linearity and regression, see e.g. Williams [Wil], Ch. 8, or [BinFr].

Acknowledgements. This work arises out of collaboration with Akihiko Inoue of Hiroshima University and Yukio Kasahara of Hokkaido University. It is a pleasure to thank them both. It is also a pleasure to thank the Mathematics Departments of both universities for their hospitality, and a Japanese Government grant for financial support. I am very grateful to the referee for a thorough and constructive report, which led to many improvements.

References

[Ach] N. I. Achieser, *Theory of approximation*, Frederick Ungar, New York, 1956.

- [BaiCr] A. Bain and D. Crisan, *Fundamentals of stochastic filtering*, Springer, 2009.
- [BarN-S] O. E. Barndorff-Nielsen and G. Schou, On the parametrization of autoregressive models by partial autocorrelation. *J. Multivariate Analysis* **3** (1973), 408-419.
- [BasW] E. L. Basor and H. Widom, On a Toeplitz determinant identity of Borodin and Okounkov. *Integral Equations Operator Theory* **37** (2000), 397-401.
- [Bax1] G. Baxter, A convergence equivalence related to polynomials orthogonal on the unit circle. *Trans. Amer. Math. Soc.* **99** (1961), 471-487.
- [Bax2] G. Baxter, An asymptotic result for the finite predictor. *Math. Scand.* **10** (1962), 137-144.
- [Bax3] G. Baxter, A norm inequality for a "finite-section" Wiener-Hopf equation. *Illinois J. Math.* **7** (1963), 97-103.
- [BekXu] T. Bekjan and Q. Xu, Riesz and Szegő type factorizations for non-commutative Hardy spaces. *J. Operator Theory* **62** (2009), 215-231.
- [Ber] J. Beran, *Statistics for long-memory processes*. Chapman and Hall, London, 1994.
- [Berk] K. N. Berk, Consistent autoregressive spectral estimates. *Ann. Statist.* **2** (1974), 489-502.
- [Beu] A. Beurling, On two problems concerning linear transformations in Hilbert space, *Acta Math.* **81** (1948), 239-255 (reprinted in *The collected works of Arne Beurling*, Volumes 1,2, Birkhäuser, 1989).
- [Bin1] N. H. Bingham, Józef Marcinkiewicz: Analysis and probability. *Proc. Józef Marcinkiewicz Centenary Conference* (Poznań, 2010), Banach Centre Publications **95** (2011), 27-44.
- [Bin2] N. H. Bingham: Multivariate prediction and matrix Szegő theory. Preprint, Imperial College.
- [BinFr] N. H. Bingham and J. M. Fry, *Regression: Linear models in statistics*. Springer Undergraduate Mathematics Series, 2010.
- [BinFrKi] N. H. Bingham, J. M. Fry and R. Kiesel, Multivariate elliptic processes. *Statistica Neerlandica* **64** (2010), 352-366.
- [BinGT] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular variation*, 2nd ed., Cambridge University Press, 1989 (1st ed. 1987).
- [BinIK] N. H. Bingham, A. Inoue and Y. Kasahara, An explicit representation of Verblunsky coefficients. *Statistics and Probability Letters*, to appear; online at <http://dx.doi.org/10.1016/j.spl.2011.11.004>.
- [BLa] D. P. Blecher and L. E. Labuschagne, Applications of the Fuglede-

- Kadison determinant: Szegő's theorem and outliers for non-commutative H^p . *Trans. Amer. Math. Soc.* **360** (2008), 6131-6147.
- [Bl1] P. Bloomfield, An exponential model for the spectrum of a scalar time series. *Biometrika* **60** (1973), 217-226.
- [Bl2] P. Bloomfield, *Fourier analysis of time series: An introduction*. Wiley, 1976.
- [Bl3] P. Bloomfield, Non-singularity and asymptotic independence. *E. J. Hannan Festschrift, J. Appl. Probab.* **23A** (1986), 9-21.
- [BlJeHa] P. Bloomfield, N. P. Jewell and E. Hayashi, Characterization of completely nondeterministic stochastic processes. *Pacific J. Math.* **107** (1983), 307-317.
- [BogHeTu] B. P. Bogert, M. J. R. Healy and J. W. Tukey, The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. *Proc. Symposium on Time Series Analysis* (ed. M. Rosenblatt) Ch. 15, 209-243, Wiley, 1963.
- [BorOk] A. M. Borodin and A. Okounkov, A Fredholm determinant formula for Toeplitz determinants. *Integral Equations and Operator Theory* **37** (2000), 386-396.
- [Bot] A. Böttcher, Featured review of the Borodin-Okounkov and Basor-Widom papers. *Mathematical Reviews* 1790118/6 (2001g:47042a,b).
- [BotKaSi] A. Böttcher, A. Karlovich and B. Silbermann, Generalized Krein algebras and asymptotics of Toeplitz determinants. *Methods Funct. Anal. Topology* **13** (2007), 236-261.
- [BotSi1] A. Böttcher and B. Silbermann, *Analysis of Toeplitz operators*. Springer, 1990 (2nd ed., with A. Karlovich, 2006).
- [BotSi2] A. Böttcher and B. Silbermann, *Introduction to large truncated Toeplitz matrices*. Universitext, Springer, 1999.
- [BotW] A. Böttcher and H. Widom, Szegő via Jacobi. *Linear Algebra and Applications* **419** (2006), 656-667.
- [BouKo] J. Bourgain and G. Kozma, One cannot hear the winding number. *J. European Math. Soc.* **9** (2007), 637-658.
- [BoGePu] A. Boutet de Monvel-Berthier, V. Georgescu and R. Purice, A boundary-value problem related to the Ginzburg-Landau model. *Comm. Math. Phys.* **142** (1991), 1-23.
- [BoxJeRe] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time-series analysis. Forecasting and control* (4th ed.). Wiley, 2008.
- [Bra1] R. C. Bradley, Basic properties of strong mixing conditions. Pages 165-192 in [EbTa].

- [Bra2] R. C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2** (2005), 107-144.
- [Bra3] R. C. Bradley, *Introduction to strong mixing conditions*, Volumes 1-3. Kendrick Press, Heber City, UT, 2007.
- [dBR] L. de Branges and J. Rovnyak, *Square-summable power series*. Holt, Rinehart and Winston, New York, 1966.
- [Bre] H. Brezis, New questions related to the topological degree. *The unity of mathematics*. 137-154, Progr. Math. **244** (2006), Birkhäuser, Boston MA.
- [Bri] D. R. Brillinger, John W. Tukey's work on time series and spectrum analysis. *Ann. Statist.* **30** (2002), 1595-1918.
- [BriKri] D. R. Brillinger and P. R. Krishnaiah (ed.), *Time series in the frequency domain*. Handbook of Statistics **3**, North-Holland, 1983.
- [BroDav] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods* (2nd ed.), Springer, New York, 1991 (1st ed. 1987).
- [Cra] H. Cramér, On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.* **28B** (1942), 1-7, reprinted in *Collected Works of Harald Cramér* Volume II, 941-947, Springer, 1994.
- [CraLea] H. Cramér and R. Leadbetter, *Stationary and related stochastic processes*. Wiley, 1967.
- [Ch] K.-L. Chung, *A course in probability theory*, 3rd ed. Academic Press, 2001 (2nd ed. 1974, 1st ed. 1968).
- [CorFoSi] I. P. Cornfeld, S. V. Fomin and Ya. G. Sinai, *Ergodic theory*. Grundle. math. Wiss. **245**, Springer, 1982.
- [CovTh] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley, 1991.
- [Cox] D. R. Cox, Long-range dependence: a review. Pages 55-74 in *Statistics: An appraisal* (ed. H. A. David and H. T. David), Iowa State University Press, Ames IA, reprinted in *Selected statistical papers of Sir David Cox*, Volume 2, 379-398, Cambridge University Press, 2005.
- [Deb] L. Debowski, On processes with summable partial autocorrelations. *Statistics and Probability Letters* **77** (2007), 752-759.
- [Deg] S. Dégerine, Canonical partial autocorrelation function of a multivariate time series. *Ann. Statist.* **18** (1990), 961-971.
- [dLR] K. de Leeuw and W. Rudin, Extreme points and extremum problems in H^1 . *Pacific J. Math.* **8** (1958), 467-485.
- [Dia] P. Diaconis, G. H. Hardy and probability??? *Bull. London Math. Soc.* **34** (2002), 385-402.
- [Do] J. L. Doob, *Stochastic processes*. Wiley, 1953.

- [DouOpTa] P. Doukhan, G. Oppenheim and M. S. Taqqu (ed.), *Theory and applications of long-range dependence*. Birkhäuser, Basel, 2003.
- [DunSch] N. Dunford and J. T. Schwartz, *Linear operators, Part II: Spectral theory: Self-adjoint operators on Hilbert space*. Interscience, 1963.
- [Dur] J. Durbin, The fitting of time-series models. *Rev. Int. Statist. Inst.* **28** (1960), 233-244.
- [Du] P. L. Duren, *Theory of H^p spaces*. Academic Press, New York, 1970.
- [Dym] H. Dym, M. G. Krein's contributions to prediction theory. *Operator Theory Advances and Applications* **118** (2000), 1-15, Birkhäuser, Basel.
- [DymMcK] H. Dym and H. P. McKean, *Gaussian processes, function theory and the inverse spectral problem*. Academic Press, 1976.
- [EbTa] E. Eberlein and M. S. Taqqu (ed.), *Dependence in probability and statistics. A survey of recent results*. Birkhäuser, 1986.
- [Ell] R. S. Ellis, *Entropy, large deviations and statistical mechanics*. Grundle. math. Wiss **271**, Springer, 1985.
- [Fe] C. Fefferman, Characterizations of bounded mean oscillation. *Bull. Amer. Math. Soc.* **77** (1971), 587-588.
- [FeSt] C. Fefferman and E. M. Stein, H^p spaces of several variables. *Acta Math.* **129** (1972), 137-193.
- [Gao] J. Gao, *Non-linear time series. Semi-parametric and non-parametric models*. Monogr. Stat. Appl. Prob. **108**, Chapman and Hall, 2007.
- [Gar] J. B. Garnett, *Bounded analytic functions*. Academic Press, 1981 (Grad. Texts in Math. **236**, Springer, 2007).
- [Geo] H.-O. Georgii, *Gibbs measures and phase transitions*. Walter de Gruyter, 1988.
- [GerCa] J. S. Geronimo and K. M. Case, Scattering theory and polynomials orthogonal on the unit circle. *J. Math. Phys.* **20** (1979), 299-310.
- [Ges] F. Gesztesy, P. Deift, C. Galves, P. Perry and W. Schlag (ed.), *Spectral theory and mathematical physics: A Festschrift in honor of Barry Simon's sixtieth birthday*. Proc. Symp. Pure Math. **76** Parts 1, 2, Amer. Math. Soc., 2007.
- [GiKoSu] L. Giraitis, H. L. Koul and D. Surgailis, *Large sample inference for long memory processes*, World Scientific, 2011.
- [GolKPY] L. Golinskii, A. Kheifets, F. Peherstorfer and P. Yuditskii, Scattering theory for CMV matrices: uniqueness, Helson-Szegö and strong Szegö theorems. *Integral Equations and Operator Theory* **69** (2011), 479-508.
- [GolTo] L. Golinskii and V. Totik, Orthogonal polynomials: from Jacobi to Simon. P. 715-742 in [Ge], Part 2.

- [GolvL] G. H. Golub and C. F. van Loan, *Matrix computations*, 3rd ed., Johns Hopkins University Press, 1996 (1st ed. 1983, 2nd ed. 1989).
- [GrSz] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. University of California Press, Berkeley CA, 1958.
- [Gri1] G. R. Grimmett, *Percolation*, 2nd ed. Grundle Math. Wiss. **321**, Springer, 1999 (1st ed. 1989).
- [Gri2] G. R. Grimmett, *The random cluster model*. Grundle Math. Wiss. **333**, Springer, 2006.
- [Ha1] E. J. Hannan, *Multiple time series*. Wiley, 1970.
- [Ha2] E. J. Hannan, The Whittle likelihood and frequency estimation. Chapter 15 (p. 205-212) in [Kel].
- [HaKR] E. J. Hannan, P. K. Krishnaiah and M. M. Rao, *Time series in the time domain*. Handbook of Statistics **5**, North-Holland, 1985.
- [He] H. Helson, *Harmonic analysis*, 2nd ed., Hindustan Book Agency, 1995.
- [HeSa] H. Helson and D. Sarason, Past and future. *Math. Scand* **21** (1967), 5-16.
- [HeSz] H. Helson and G. Szegö, A problem in prediction theory. *Acta Mat. Pura Appl.* **51** (1960), 107-138.
- [Ho] K. Hoffman, *Banach spaces of analytic functions*, Prentice-Hall, Englewood Cliffs NJ, 1962.
- [Hos] J. R. Hosking, Fractional differencing. *Biometrika* **68** (1981), 165-176.
- [HuMuWe] R. A. Hunt, B. Muckenhoupt and R. L. Wheeden, Weighted norm inequalities for the conjugate function and Hilbert transform. *Trans. Amer. Math. Soc.* **176** (1973), 227-151.
- [IbLi] I. A. Ibragimov and Yu. V. Linnik, *Independent and stationary sequences of random variables*. Wolters-Noordhoff, 1971.
- [IbRo] I. A. Ibragimov and Yu. A. Rozanov, *Gaussian random processes*. Springer, 1978.
- [In1] A. Inoue, Asymptotics for the partial autocorrelation function of a stationary process. *J. Analyse Math.* **81** (2000), 65-109.
- [In2] A. Inoue, Asymptotic behaviour for partial autocorrelation functions of fractional ARIMA processes. *Ann. Appl. Probab.* **12** (2002), 1471-1491.
- [In3] A. Inoue, AR and MA representations of partial autocorrelation functions, with applications. *Prob. Th. Rel. Fields* **140** (2008), 523-551.
- [InKa1] A. Inoue and Y. Kasahara, Partial autocorrelation functions of the fractional ARIMA processes. *J. Multivariate Analysis* **89** (2004), 135-147.
- [InKa2] A. Inoue and Y. Kasahara, Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34** (2006), 973-993.

- [Jan] S. Janson, *Gaussian Hilbert spaces*. Cambridge Tracts in Math. **129**, Cambridge University Press, 1997.
- [JeBl] N. P. Jewell and P. Bloomfield, Canonical correlations of past and future for time series: definitions and theory. *Ann. Statist.* **11** (1983), 837-847.
- [JeBlBa] N. P. Jewell, P. Bloomfield and F. C. Bartmann, Canonical correlations of past and future for time series: bounds and computation. *Ann. Statist.* **11** (1983), 848-855.
- [Kac] M. Kac, Toeplitz matrices, transition kernels and a related problem in probability theory. *Duke Math. J.* **21** (1954), 501-509.
- [KahSa] J.-P. Kahane and R. Salem, *Ensembles parfaits et séries trigonométriques*, 2nd ed. Hermann, Paris, 1994.
- [Kak] Y. Kakihara, The Kolmogorov isomorphism theorem and extensions to some non-stationary processes. *Stochastic processes: Theory and methods* (ed. D. N. Shanbhag and C. R. Rao), *Handbook of Statistics* **19**, North-Holland, 2001, 443-470.
- [KanSch] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press, 1997 (2nd ed. 2004).
- [KaBi] Y. Kasahara and N. H. Bingham, Verblunsky coefficients and Nehari sequences. Preprint, Hokkaido University.
- [KatSeTe] D. Kateb, A. Seghier and G. Teyssi re, Prediction, orthogonal polynomials and Toeplitz matrices. A fast and reliable approach to the Durbin-Levinson algorithm. Pages 239-261 in [TK].
- [Kel] F. P. Kelly (ed.), *Probability, statistics and optimization. A tribute to Peter Whittle*. Wiley, 1994.
- [KenSt] M. G. Kendall and A. Stuart, *The advanced theory of statistics*. Charles Griffin. Volume 1 (4th ed., 1977), Vol. 2 (3rd ed, 1973), vol. 3 (3rd ed., 1976).
- [KokTa] P. S. Kokoszka and M. S. Taqqu, Can one use the Durbin-Levinson algorithm to generate infinite-variance fractional ARIMA time series? *J. Time Series Analysis* **22** (2001), 317-337.
- [Kol] A. N. Kolmogorov, Stationary sequences in Hilbert space. *Bull. Moskov. Gos. Univ. Mat.* **2** (1941), 1-40 (in Russian; reprinted, *Selected works of A. N. Kolmogorov, Vol. 2: Theory of probability and mathematical statistics*, Nauka, Moskva, 1986, 215-255).
- [Koo1] P. Koosis, *Introduction to H^p spaces*, 2nd ed. Cambridge Tracts Math. **115**, Cambridge Univ. Press, 1998 (1st ed. 1980).
- [Koo2] P. Koosis, *The logarithmic integral*, I, 2nd ed., Cambridge Univ.

- Press, 1998 (1st ed. 1988), II, Cambridge Univ. Press, 1992.
- [Kr] M. G. Krein, On some new Banach algebras and Wiener-Lévy type theorems for Fourier series and integrals. *Amer. Math. Soc. Translations* (2) **93** (1970), 177-199 (Russian original: *Mat. Issled.* **1** (1966), 163-288).
- [Lev] N. Levinson, The Wiener (RMS) error criterion in filter design and prediction. *J. Math. Phys. MIT* **25** (1947), 261-278.
- [LevMcK] N. Levinson and H. P. McKean, Weighted trigonometrical approximation on R^1 with application to the germ field of a stationary Gaussian noise. *Acta Math.* **112** (1964), 99-143.
- [Li] L. M. Li, Some notes on mutual information between past and future. *J. Time Series Analysis* **27** (2006), 309-322.
- [LiXi] L. M. Li and Z. Xie, Model selection and order determination for time series by information between the past and the future. *J. Time Series Analysis* **17** (1996), 65-84.
- [LuZhKi] R. Lund, Y. Zhao and P. C. Kiessler, Shapes of stationary autocovariances. *J. Applied Probability* **43** (2006), 1186-1193.
- [Ly1] R. Lyons, Characterizations of measures whose Fourier-Stieltjes transforms vanish at infinity. *Bull. Amer. Math. Soc.* **10** (1984), 93-96.
- [Ly2] R. Lyons, Fourier-Stieltjes coefficients and asymptotic distribution modulo 1. *Ann. Math.* **122** (1985), 155-170.
- [Ly3] R. Lyons, Seventy years of Rajchman measures. *J. Fourier Anal. Appl.*, Kahane Special Issue (1995), 363-377.
- [MakWe] A. Makagon and A. Weron, q -variate minimal stationary processes. *Studia Math.* **59** (1976), 41-52.
- [MatNeTo] A. Máté, P. Nevai and V. Totik, Asymptotics for the ratio of leading coefficients of orthogonal polynomials on the unit circle. *Constructive Approximation* **1** (1985), 63-69.
- [McCW] B. M. McCoy and T. T. Wu, *The two-dimensional Ising model*. Harvard Univ. Press, Cambridge MA, 1973.
- [McC] P. McCullagh, John Wilder Tukey, 1915-2000. *Biographical Memoirs of Fellows of the Royal Society* **49** (2003), 537-555.
- [McLZ] A. I. McLeod and Y. Zhang, Partial autocorrelation parametrization for subset regression. *J. Time Series Analysis* **27** (2006), 599-612.
- [Me] Y. Meyer, *Wavelets and operators*. Cambridge Univ. Press, 1992.
- [MeCo] Y. Meyer and R. Coifman, *Wavelets. Calderón-Zygmund and multilinear operators*. Cambridge Univ. Press, 1997.
- [Nak1] T. Nakazi, Exposed points and extremal problems in H^1 . *J. Functional Analysis* **53** (1983), 224-230.

- [Nak2] T. Nakazi, Exposed points and extremal problems in H^1 , II. *Tohoku Math. J.* **37** (1985), 265-269.
- [vN] J. von Neumann, Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. *Math. Ann.* **102**, 49-131 (*Collected Works* II.1).
- [Nik1] N. K. Nikolskii, *Treatise on the shift operator: Spectral function theory*. Grundle. math. Wiss. **273**, Springer, 1986.
- [Nik2] N. K. Nikolskii, *Operators, functions and systems: an easy reading. Volume 1: Hardy, Hankel and Toeplitz; Volume 2: Model operators and systems*. Math. Surveys and Monographs **92, 93**, Amer. Math. Soc., 2002.
- [OpSc] A. V. Oppenheim and R. W. Schaffer, *Discrete signal processing*. Prentice-Hall, 1989.
- [Pel] V. V. Peller, *Hankel operators and their applications*. Springer, 2003.
- [PoSz] G. Pólya and G. Szegő, *Problems and theorems in analysis*, I, II. Classics in Math., Springer, 1998 (transl. 4th German ed., 1970; 1st ed. 1925).
- [Pou] M. Pourahmadi, *Foundations of time series analysis and prediction theory*. Wiley, 2001.
- [Rak] E. A. Rakhmanov, On the asymptotics of the ratios of orthogonal polynomials, II, *Math. USSR Sb.* **58** (1983), 105-117.
- [Ram] F. L. Ramsey, Characterization of the partial autocorrelation function. *Ann. Statist.* **2** (1974), 1296-1301.
- [Rao] M. M. Rao, Harmonizable, Cramér and Karhunen classes of processes. Ch. 10 (p.276-310) in [HaKR].
- [Rob] P. M. Robinson (ed.), *Time series with long memory*. Advanced Texts in Econometrics, Oxford University Press, 2003.
- [RoRo] M. Rosenblum and J. Rovnyak, *Hardy classes and operator theory*, Dover, New York, 1997 (1st ed. Oxford University Press, 1985).
- [Roz] Yu. A. Rozanov, *Stationary random processes*. Holden-Day, 1967.
- [Ru] W. Rudin, *Real and complex analysis*, 2nd ed. McGraw-Hill, 1974 (1st ed. 1966).
- [Sa1] D. Sarason, *Function theory on the unit circle*. Virginia Polytechnic Institute and State University, Blacksburg VA, 1979.
- [Sa2] D. Sarason, An addendum to "Past and future", *Math. Scand.* **30** (1972), 62-64.
- [Sa3] D. Sarason, Functions of vanishing mean oscillation, *Trans. Amer. Math. Soc.* **207** (1975), 391-405.
- [Si1] B. Simon, *The statistical mechanics of lattice gases, Volume 1*. Princeton University Press, 1993.
- [Si2] B. Simon, The Golinskii-Ibragimov method and a theorem of Damanik-

- Killip. *Int. Math. Res. Notes* (2003), 1973-1986.
- [Si3] B. Simon, OPUC on one foot. *Bull. Amer. Math. Soc.* **42** (2005), 431-460.
- [Si4] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory*. AMS Colloquium Publications 54.1, American Math. Soc., Providence RI, 2005.
- [Si5] B. Simon, *Orthogonal polynomials on the unit circle. Part 2: Spectral theory*. AMS Colloquium Publications 54.2, American Math. Soc., Providence RI, 2005.
- [Si6] B. Simon, The sharp form of the strong Szegő theorem. *Contemporary Math.* **387** (2005), 253-275, AMS, Providence RI.
- [Si7] B. Simon, Meromorphic Szegő functions and asymptotic series for Verblunsky coefficients. *Acta Math.* **195** (2005), 267-285.
- [Si8] B. Simon, Ed Nelson's work in quantum theory. *Diffusion, quantum theory and radically elementary mathematics* (ed. W. G. Faris), *Math. Notes* **47** (2006), 75-93.
- [Si9] B. Simon, *Szegő's theorem and its descendants: Spectral theory for L^2 perturbations of orthogonal polynomials*. Princeton University Press, 2011.
- [Si10] B. Simon, *Convexity: An analytic viewpoint*. Cambridge Tracts in Math. **187**, Cambridge University Press, 2011.
- [Sz1] G. Szegő, Ein Grenzwertsatz über die Toeplitzschen Determinanten einer reellen positiven Funktion. *Math. Ann.* **76** (1915), 490-503.
- [Sz2] G. Szegő, Beiträge zur Theorie der Toeplitzschen Formen. *Math. Z.* **6** (1920), 167-202.
- [Sz3] G. Szegő, Beiträge zur Theorie der Toeplitzschen Formen, II. *Math. Z.* **9** (1921), 167-190.
- [Sz4] G. Szegő, *Orthogonal polynomials*. AMS Colloquium Publications 23, American Math. Soc., Providence RI, 1939.
- [Sz5] G. Szegő, On certain Hermitian forms associated with the Fourier series of a positive function. *Festschrift Marcel Riesz* 222-238, Lund, 1952.
- [SzNF] B. Sz.-Nagy and C. Foias, *Harmonic analysis of operators on Hilbert space*, North-Holland, 1970 (2nd ed., with H. Bercovici and L. Kérchy, Springer Universitext, 2010).
- [TeKi] G. Teyssière and A. P. Kirman (ed.), *Long memory in economics*. Springer, 2007.
- [Tor] A. Torchinsky, *Real-variable methods in harmonic analysis*, Dover, 2004 (Academic Press, 1981).
- [TrVo1] S. Treil and A. Volberg, Wavelets and the angle between past and

- future. *J. Functional analysis* **143** (1997), 269-308.
- [TrVo2] S. Treil and A. Volberg, A simple proof of the Hunt-Muckenhoupt-Wheeden theorem. Preprint, 1997.
- [Ts] B. Tsirelson, Spectral densities describing off-white noises. *Ann. Inst. H. Poincaré Prob. Stat.* **38** (2002), 1059-1069.
- [V1] On positive harmonic functions. A contribution to the algebra of Fourier series. *Proc. London Math. Soc.* **38** (1935), 125-157.
- [V2] On positive harmonic functions (second paper). *Proc. London Math. Soc.* **40** (1936), 290-320.
- [Wh] P. Whittle, *Hypothesis testing in time series analysis*. Almqvist and Wiksell, Uppsala, 1951.
- [Wi1] N. Wiener, Generalized harmonic analysis. *Acta Math.* **55** (1930), 117-258 (reprinted in *Generalized harmonic analysis and Tauberian theorems*, MIT Press, Cambridge MA, 1986, and *Collected Works, Volume II: Generalized harmonic analysis and Tauberian theory; classical harmonic and complex analysis* (ed. P. Masani), MIT Press, Cambridge MA, 1979).
- [Wi2] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series. With engineering applications*. MIT Press/Wiley, 1949.
- [Wi3] N. Wiener, *Collected Works, Volume III: The Hopf-Wiener integral equation; prediction and filtering; quantum mechanics and relativity; miscellaneous mathematical papers* (ed. P. Masani), MIT Press, Cambridge MA, 1981.
- [Wil] D. Williams, *Weighing the odds*. Cambridge University Press, 2001.
- [Wo] H. Wold, *A study in the analysis of stationary time series*. Almqvist and Wiksell, Uppsala, 1938 (2nd ed., appendix by Peter Whittle, 1954).
- [Z1] A. Zygmund, Sur les fonctions conjuguées. *Fund. Math.* **13** (1929), 284-303; corr. *Fund. Math.* **18** (1932), 312 (reprinted in [Z3], vol 1).
- [Z2] A. Zygmund, *Trigonometric series*, Volumes 1,2, Cambridge University Press, 1968.
- [Z3] A. Zygmund, *Selected papers of Antoni Zygmund* (ed. A. Hulanicki, P. Wojtaszczyk and W. Zelasko), Volumes 1-3, Kluwer, Dordrecht, 1989.

N. H. Bingham, Mathematics Department, Imperial College London, London SW7 2AZ, UK nick.bingham@btinternet.com n.bingham@ic.ac.uk